TWICE-RAMANUJAN SPARSIFIERS[*]

JOSHUA BATSON[†], DANIEL A. SPIELMAN[‡], AND NIKHIL SRIVASTAVA[§]

Abstract. We prove that every graph has a spectral sparsifier with a number of edges linear in its number of vertices. As linear-sized spectral sparsifiers of complete graphs are expanders, our sparsifiers of arbitrary graphs can be viewed as generalizations of expander graphs. In particular, we prove that for every $d > 1$ and every undirected, weighted graph $G = (V, E, w)$ on n vertices, there exists a weighted graph $H = (V, F, \tilde{w})$ with at most $\lceil d(n-1) \rceil$ edges such that for every $x \in \mathbb{R}^V$, $x^T L_G x \le x^T L_H x \le \left(\frac{d+1+2\sqrt{d}}{d+1-2\sqrt{d}} \right) \cdot x^T L_G x$, where L_G and L_H are the Laplacian matrices of G and H, respectively. Thus, H approximates G spectrally at least as well as a Ramanujan expander with $dn/2$ edges approximates the complete graph. We give an elementary deterministic polynomial time algorithm for constructing H.

Key words. graph sparsification, spectral graph theory, spectral sparsification, rank-one updates, barrier method, paving conjecture, sparse quadratic forms

AMS subject classifications. Primary, 68R10; Secondary, 15B48

DOI. 10.1137/090772873

1. Introduction. A sparsifier of a graph $G = (V, E, w)$ is a sparse graph H that is similar to G in some useful way. Many notions of similarity have been considered. For example, Chew's [6] spanners have the property that the distance between every pair of vertices in H is approximately the same as in G. Benczúr and Karger's [3] cut sparsifiers have the property that the weight of the boundary of every set of vertices is approximately the same in G as in H. We consider the spectral notion of similarity introduced by Spielman and Teng [17, 19]: we say that H is a κ-approximation of G if for all $x \in \mathbb{R}^V$,

$$(1) \qquad x^T L_G x \le x^T L_H x \le \kappa \cdot x^T L_G x,$$

where L_G and L_H are the Laplacian matrices of G and H. We recall that

$$x^T L_G x = \sum_{(u,v) \in E} w_{u,v} (x_u - x_v)^2,$$

where $w_{u,v}$ is the weight of edge (u, v) in G. By considering vectors x that are the characteristic vectors of sets, one can see that condition (1) is strictly stronger than the cut condition of Benczúr and Karger.

In the case where G is the complete graph, excellent spectral sparsifiers are supplied by *Ramanujan graphs* [12, 13]. These are d-regular graphs H all of whose

[†]Department of Mathematics, MIT, Cambridge, MA 02139 (joshuab@mac.com). Work on this paper was performed while this author was at Yale College.

[‡]Program in Applied Mathematics and Department of Computer Science, Yale University, New Haven, CT 06520 (spielman@cs.yale.edu).

[§]Department of Computer Science, Yale University, New Haven, CT 06520 (nikhils@math.ias.edu).

nonzero Laplacian eigenvalues lie between $d - 2\sqrt{d-1}$ and $d + 2\sqrt{d-1}$. Thus, if we take a Ramanujan graph on n vertices and multiply the weight of every edge by $n/(d - 2\sqrt{d-1})$, we obtain a graph that κ-approximates the complete graph for

$$\kappa = \frac{d + 2\sqrt{d-1}}{d - 2\sqrt{d-1}}.$$

In this paper, we prove that every graph can be approximated at least this well[1] by a graph with only twice as many edges as the Ramanujan graph (as a d-regular graph has $dn/2$ edges).

THEOREM 1.1. *For every $d > 1$, every undirected weighted graph $G = (V, E, w)$ on n vertices contains a weighted subgraph $H = (V, F, \tilde{w})$ with $\lceil d(n-1) \rceil$ edges (i.e., average degree at most $2d$) that satisfies*

$$x^T L_G x \leq x^T L_H x \leq \left(\frac{d + 1 + 2\sqrt{d}}{d + 1 - 2\sqrt{d}} \right) \cdot x^T L_G x \qquad \forall x \in \mathbb{R}^V.$$

Our proof provides a deterministic greedy algorithm for computing the graph H in time $O(dn^3 m)$.

We remark that while the edges of H are a subset of the edges of G, the weights of edges in H and G will typically be different. In fact, there exist unweighted graphs G for which every good spectral sparsifier H must contain edges of widely varying weights [19].

1.1. Expanders: Sparsifiers of the complete graph. In the case when G is a complete graph, our construction produces expanders. However, these expanders are slightly unusual in that their edges have weights, they may be irregular, and the weighted degrees of vertices can vary slightly. This may lead one to ask whether they should really be considered expanders. In section 4 we argue that they should be.

As the graphs we produce are irregular and weighted, it is also not immediately clear that we should be comparing κ with the Ramanujan bound of

$$(2) \qquad \frac{d + 2\sqrt{d-1}}{d - 2\sqrt{d-1}} = 1 + \frac{4}{\sqrt{d}} + O(1/d).$$

It is known[2] that no d-regular graph of uniform weight can κ-approximate a complete graph for κ asymptotically better than (2) [14]. While we believe that no graph of *average* degree d can be a κ-approximation of a complete graph for κ asymptotically better than (2), we are unable to show this at the moment and prove instead the weaker claim that no such graph can achieve κ less than

$$1 + \frac{2}{\sqrt{d}} - O\left(\frac{\sqrt{d}}{n} \right).$$

---

[1]Strictly speaking, our approximation constant is only better than the Ramanujan bound $\kappa = \frac{d+2\sqrt{d-1}}{d-2\sqrt{d-1}}$ in the regime $d \geq \frac{1+\sqrt{5}}{2}$. This includes the actual Ramanujan graphs, for which d is an integer greater than 2.

[2]While lower bounds on the spectral gap of d-regular graphs focus on showing that the second smallest eigenvalue is asymptotically at most $d - 2\sqrt{d-1}$, the same proofs by test functions can be used to show that the largest eigenvalue is asymptotically at least $d + 2\sqrt{d-1}$.

1.2. Prior work. Spielman and Teng [17, 19] introduced the notion of sparsification that we consider, and proved that $(1 + \epsilon)$-approximations with $\widetilde{O}\left(n/\epsilon^2\right)$ edges could be constructed in $\widetilde{O}\left(m\right)$ time. They used these sparsifiers to obtain a nearly linear time algorithm for solving diagonally dominant systems of linear equations [17, 18].

Spielman and Teng were inspired by the notion of sparsification introduced by Benczúr and Karger [3] for cut problems, which only required inequality (1) to hold for all $x \in \{0, 1\}^V$. Benczúr and Karger showed how to construct graphs H meeting this guarantee with $O\left(n \log n/\epsilon^2\right)$ edges in $O\left(m \log^3 n\right)$ time; their cut sparsifiers have been used to obtain faster algorithms for cut problems [3, 11].

Spielman and Srivastava [16] proved the existence of spectral sparsifiers with $O\left(n \log n/\epsilon^2\right)$ edges and showed how to construct them in $\widetilde{O}\left(m\right)$ time. They conjectured that it should be possible to find such sparsifiers with only $O\left(n/\epsilon^2\right)$ edges. We affirmatively resolve this conjecture.

Recently, partial progress was made toward this conjecture by Goyal, Rademacher, and Vempala [9], who showed how to find graphs H with only $2n$ edges that $O\left(\log n\right)$-approximate bounded degree graphs G under the cut notion of Benczúr and Karger.

We remark that all of these constructions were randomized. Ours is the first deterministic algorithm to achieve the guarantees of any of these papers.

2. Preliminaries.

2.1. The incidence matrix and the Laplacian. Let $G = (V, E, w)$ be a connected weighted undirected graph with n vertices and m edges and edge weights $w_e > 0$. If we orient the edges of G arbitrarily, we can write its Laplacian as $L = B^T W B$, where $B_{m \times n}$ is the *signed edge-vertex incidence matrix* given by

$$B(e, v) = \begin{cases} 1 & \text{if } v \text{ is } e\text{'s head,} \\ -1 & \text{if } v \text{ is } e\text{'s tail,} \\ 0 & \text{otherwise,} \end{cases}$$

and $W_{m \times m}$ is the diagonal matrix with $W(e, e) = w_e$. It is immediate that L is positive semidefinite since

$$x^T L x = x^T B^T W B x = \|W^{1/2} B x\|_2^2$$
$$= \sum_{(u,v) \in E} w_{u,v} (x_u - x_v)^2 \geq 0 \quad \text{for every } x \in \mathbb{R}^n$$

and that G is connected if and only if $\ker(L) = \ker(W^{1/2} B) = \text{span}(\mathbf{1})$.

2.2. The pseudoinverse. Since L is symmetric we can diagonalize it and write

$$L = \sum_{i=1}^{n-1} \lambda_i u_i u_i^T,$$

where $\lambda_1, \ldots, \lambda_{n-1}$ are the nonzero eigenvalues of L and $u_1, \ldots, u_{n-1}$ are a corresponding set of orthonormal eigenvectors. The *Moore–Penrose pseudoinverse* of L is then defined as

$$L^+ = \sum_{i=1}^{n-1} \frac{1}{\lambda_i} u_i u_i^T.$$

Notice that $\ker(L) = \ker(L^+)$ and that

$$LL^+ = L^+L = \sum_{i=1}^{n-1} u_i u_i^T,$$

which is simply the projection onto the span of the nonzero eigenvectors of L (which are also the eigenvectors of L^+). Thus, $LL^+ = L^+L$ is the identity on $\mathrm{im}(L) = \ker(L)^\perp$.

2.3. Formulas for rank-one updates. We use the following well-known theorem from linear algebra, which describes the behavior of the inverse of a matrix under rank-one updates (see [8, section 2.1.3]).

LEMMA 2.1 (Sherman–Morrison formula). *If A is a nonsingular $n \times n$ matrix and $\mathbf{v}$ is a vector, then*

$$(A + \mathbf{v}\mathbf{v}^T)^{-1} = A^{-1} - \frac{A^{-1}\mathbf{v}\mathbf{v}^T A^{-1}}{1 + \mathbf{v}^T A^{-1}\mathbf{v}}.$$

There is a related formula describing the change in the *determinant* of a matrix under the same update.

LEMMA 2.2 (matrix determinant lemma). *If A is nonsingular and $\mathbf{v}$ is a vector, then*

$$\det(A + \mathbf{v}\mathbf{v}^T) = \det(A)(1 + \mathbf{v}^T A^{-1}\mathbf{v}).$$

3. The main result. At the heart of this work is the following purely linear algebraic theorem. We use the notation $A \preceq B$ to mean that $B - A$ is positive semidefinite and $\mathbf{id}_S$ to denote the identity operator on a vector space S.

THEOREM 3.1. *Suppose that $d > 1$ and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ are vectors in $\mathbb{R}^n$ with*

$$\sum_{i \le m} \mathbf{v}_i \mathbf{v}_i^T = \mathbf{id}_{\mathbb{R}^n}.$$

Then there exist scalars $s_i \ge 0$ with $|\{i : s_i \ne 0\}| \le dn$ so that

$$\mathbf{id}_{\mathbb{R}^n} \preceq \sum_{i \le m} s_i \mathbf{v}_i \mathbf{v}_i^T \preceq \left(\frac{d+1+2\sqrt{d}}{d+1-2\sqrt{d}}\right) \mathbf{id}_{\mathbb{R}^n}.$$

The sparsification result for graphs follows quickly from this theorem as shown below.

Proof of Theorem 1.1. Assume without loss of generality that G is connected. Write $L_G = B^T W B$ as in section 2.1 and fix $d > 1$. Restrict attention to $\mathrm{im}(L_G) \cong \mathbb{R}^{n-1}$ and apply Theorem 3.1 to the columns $\{\mathbf{v}_i\}_{i \le m}$ of

$$V_{n \times m} = (L_G^+)^{\frac{1}{2}} B^T W^{\frac{1}{2}},$$

which are indexed by the edges of G and satisfy

$$\sum_{i \le m} \mathbf{v}_i \mathbf{v}_i^T = V V^T = (L_G^+)^{\frac{1}{2}} B^T W B (L_G^+)^{\frac{1}{2}}$$

$$= (L_G^+)^{\frac{1}{2}} L_G (L_G^+)^{\frac{1}{2}} = \mathbf{id}_{\mathrm{im}(L_G)}.$$

Write the scalars $s_i \geq 0$ guaranteed by the theorem in the $m \times m$ diagonal matrix $S(i,i) = s_i$ and set $L_H = B^T W^{\frac{1}{2}} S W^{\frac{1}{2}} B$. Then L_H is the Laplacian of the subgraph H of G with edge weights $\{\tilde{w}_i = w_i s_i\}_{i \in E}$, and H has at most $d(n-1)$ edges since at most that many of the s_i are nonzero. Also,

$$\mathbf{id}_{\mathrm{im}(L_G)} \preceq \sum_{i \leq m} s_i \mathbf{v}_i \mathbf{v}_i^T = VSV^T \preceq \kappa \cdot \mathbf{id}_{\mathrm{im}(L_G)} \qquad \text{for } \kappa = \frac{d+1+2\sqrt{d}}{d+1-2\sqrt{d}}.$$

By the Courant–Fischer theorem, this is equivalent to

$$1 \leq \frac{y^T V S V^T y}{y^T y} \leq \kappa \qquad \forall y \in \mathrm{im}((L_G)^{\frac{1}{2}}) = \mathrm{im}(L_G)$$

$$\Longleftrightarrow 1 \leq \frac{y^T (L_G^+)^{\frac{1}{2}} L_H (L_G^+)^{\frac{1}{2}} y}{y^T y} \leq \kappa \qquad \forall y \in \mathrm{im}((L_G)^{\frac{1}{2}})$$

$$\Longleftrightarrow 1 \leq \frac{x^T L_G^{\frac{1}{2}} (L_G^+)^{\frac{1}{2}} L_H (L_G^+)^{\frac{1}{2}} L_G^{\frac{1}{2}} x}{x^T L_G^{\frac{1}{2}} L_G^{\frac{1}{2}} x} \leq \kappa \qquad \forall x \perp \mathbf{1}$$

$$\Longleftrightarrow 1 \leq \frac{x^T L_H x}{x^T L_G x} \leq \kappa \qquad \forall x \perp \mathbf{1},$$

as desired.   □

It is worth mentioning that the above reduction is essentially the same as the one in [16]. In that paper, the authors consider the symmetric projection matrix $\Pi = B L_G^+ B^T$ whose columns $\{\Pi_e\}_{e \in E}$ correspond to the edges of G. They show, by a concentration lemma of Rudelson [15], that randomly sampling $O(n \log n)$ of the columns with probabilities proportional to $\|\Pi_e\|^2 = w_e \mathsf{R}_{\mathrm{eff}}(e)$ (where $\mathsf{R}_{\mathrm{eff}}$ is the effective resistance) gives a matrix $\tilde{\Pi}$ that approximates Π in the spectral norm and corresponds to a graph sparsifier, with high probability. In this paper, we do essentially the same thing with two modifications: we eliminate Π in order to simplify notation, since we are no longer following the intuition of sampling by effective resistances and, instead of Rudelson's sampling lemma, we use Theorem 3.1 to *deterministically* select $O(n)$ edges (equivalently, columns of Π).

The rest of this section is devoted to proving Theorem 3.1. The proof is constructive and yields a deterministic polynomial time algorithm for finding the scalars s_i, which can then be used to sparsify graphs, as advertised.

Given vectors $\{\mathbf{v}_i\}$, our goal is to choose a small set of coefficients s_i so that $A = \sum_i s_i \mathbf{v}_i \mathbf{v}_i^T$ is well-conditioned. We will build the matrix A in steps, starting with $A = 0$ and adding one vector $s_i \mathbf{v}_i \mathbf{v}_i^T$ at a time. Before beginning the proof, it will be instructive to study how the eigenvalues and characteristic polynomial of a matrix evolve upon the addition of a vector. This discussion should provide some intuition for the structure of the proof and demystify the origin of the "twice-Ramanujan" number $\frac{d+1+2\sqrt{d}}{d+1-2\sqrt{d}}$, which appears in our final result.

3.1. Intuition for the proof. It is well known that the eigenvalues of $A + \mathbf{v}\mathbf{v}^T$ interlace those of A. In fact, the new eigenvalues can be determined exactly by looking at the characteristic polynomial of $A + \mathbf{v}\mathbf{v}^T$, which is computed using Lemma 2.2 as follows:

$$p_{A+\mathbf{v}\mathbf{v}^T}(x) = \det(xI - A - \mathbf{v}\mathbf{v}^T) = p_A(x) \left(1 - \sum_j \frac{\langle \mathbf{v}, u_j \rangle^2}{x - \lambda_i} \right),$$

where λ_i are the eigenvalues of A and u_j are the corresponding eigenvectors. The polynomial $p_{A+\mathbf{v}\mathbf{v}^T}(x)$ has two kinds of zeros λ:

1. Those for which $p_A(\lambda) = 0$. These are equal to the eigenvalues λ_j of A for which the added vector $\mathbf{v}$ is orthogonal to the corresponding eigenvector u_j, and which therefore do not "move" upon adding $\mathbf{v}\mathbf{v}^T$.

2. Those for which $p_A(\lambda) \neq 0$ and

$$f(\lambda) = \left(1 - \sum_j \frac{\langle \mathbf{v}, u_j \rangle^2}{\lambda - \lambda_j}\right) = 0.$$

These are the eigenvalues which have moved and strictly interlace the old eigenvalues. The above equation immediately suggests a simple physical model (see Figure 1) which gives intuition as to where these new eigenvalues are located.
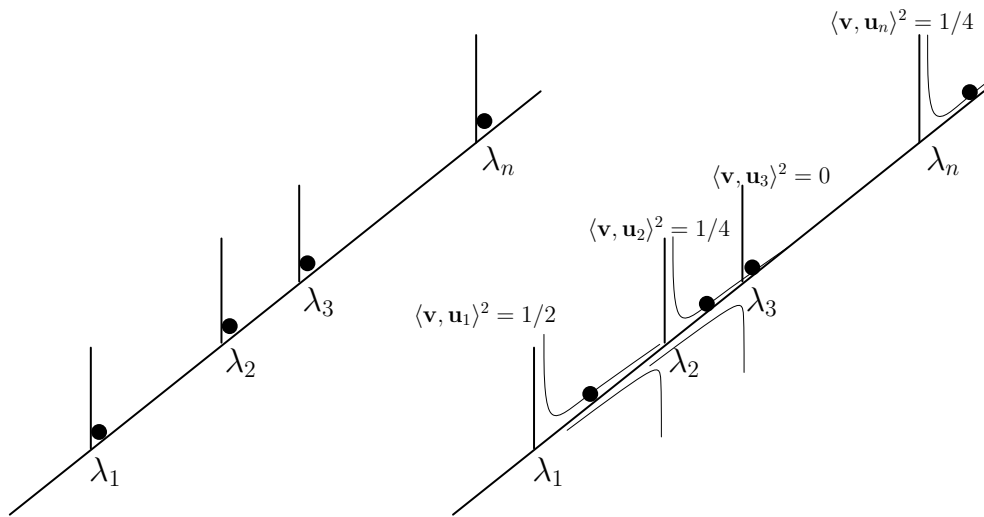

FIG. 1. *Physical model of interlacing eigenvalues.*

Physical model. We interpret the eigenvalues λ as charged particles lying on a slope. On the slope are n fixed, chargeless barriers located at the initial eigenvalues λ_j, and each particle is resting against one of the barriers under the influence of gravity. Adding the vector $\mathbf{v}\mathbf{v}^T$ corresponds to placing a charge of $\langle \mathbf{v}, u_j \rangle^2$ on the barrier corresponding to λ_j. The charges on the barriers repel those on the eigenvalues with a force that is proportional to the charge on the barrier and inversely proportional to the distance from the barrier—i.e., the force from barrier j is given by

$$\frac{\langle \mathbf{v}, u_j \rangle^2}{\lambda - \lambda_j},$$

a quantity which is positive for λ_j "below" λ, which are pushing the particle "upward," and negative otherwise. The eigenvalues move up the slope until they reach an equilibrium in which the repulsive forces from the barriers

cancel the effect of gravity, which we take to be a $+1$ in the downward direction. Thus the equilibrium condition corresponds exactly to having the total "downward pull" $f(\lambda)$ equal to zero.

With this physical model in mind, we begin to consider what happens to the eigenvalues of A when we add a *random* vector from our set $\{\mathbf{v}_i\}$. The first observation is that for any eigenvector u_j (in fact, for any vector at all), the expected projection of a randomly chosen $\mathbf{v} \in \{\mathbf{v}_i\}_{i \leq m}$ is

$$\mathbb{E}_{\mathbf{v}} \langle \mathbf{v}, u_j \rangle^2 = \frac{1}{m} \sum_i \langle \mathbf{v}_i, u_j \rangle^2 = \frac{1}{m} u_j^T \left(\sum_i \mathbf{v}_i \mathbf{v}_i^T \right) u_j = \frac{\|u_j\|^2}{m} = \frac{1}{m}.$$

Of course, this does not mean that there is any single vector $\mathbf{v}_i$ in our set that realizes this "expected behavior" of equal projections on the eigenvectors. But if we were to add such a vector[3] in our physical model, we would add equal charges of $1/m$ to each of the barriers, and we would expect all of the eigenvalues of A to drift forward "steadily." In fact, one might expect that after sufficiently many iterations of this process, the eigenvalues would all march forward together, with no eigenvalue too far ahead or too far behind, and we would end up in a position where $\lambda_{max}/\lambda_{min}$ is bounded.

In fact, this intuition turns out to be correct. Adding a vector with equal projections changes the characteristic polynomial in the following manner:

$$p_{A+\mathbf{v}_{\mathsf{avg}}\mathbf{v}_{\mathsf{avg}}^T}(x) = p_A(x) \left(1 - \sum_j \frac{1/m}{x - \lambda_j} \right) = p_A(x) - (1/m)p_A'(x),$$

since $p_A'(x) = \sum_j \prod_{i \neq j}(x - \lambda_i)$. If we start with $A = 0$, which has characteristic polynomial $p_0(x) = x^n$, then after k iterations of this process we obtain the polynomial

$$p_k(x) = (I - (1/m)D)^k x^n,$$

where D is the derivative with respect to x. Fortunately, iterating the operator $(I - \alpha D)$ for any $\alpha > 0$ generates a standard family of orthogonal polynomials—the *associated Laguerre polynomials* [7]. These polynomials are very well studied and the locations of their zeros are known; in particular, after $k = dn$ iterations the ratio of the largest to the smallest zero is known [7] to be

$$\frac{d + 1 + 2\sqrt{d}}{d + 1 - 2\sqrt{d}},$$

which is exactly what we want.

To prove the theorem, we will show that we can choose a sequence of actual vectors that realizes the expected behavior (i.e., the behavior of repeatedly adding $\mathbf{v}_{\mathsf{avg}}$) as long as we are allowed to add arbitrary fractional amounts of the $\mathbf{v}_i \mathbf{v}_i^T$ via the weights $s_i \geq 0$. We will control the eigenvalues of our matrix by maintaining two barriers as in the physical model and keeping the eigenvalues between them. The lower barrier will "repel" the eigenvalues forward; the upper one will make sure they do not go too far. The barriers will move forward at a steady pace. By maintaining that the total "repulsion" at every step of this process is bounded, we will be able to guarantee that there is always some multiple of a vector to add that allows us to continue the process.

---

[3]For concreteness, we remark that this "average" vector would be precisely $\mathbf{v}_{\mathsf{avg}} = \frac{1}{\sqrt{m}} \sum_j u_j$.

3.2. Proof by barrier functions. We begin by defining two "barrier" potential functions which measure the quality of the eigenvalues of a matrix. These potential functions are inspired by the inverse law of repulsion in the physical model discussed in the last section.

DEFINITION 3.2. *For $u, l \in \mathbb{R}$ and A, a symmetric matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, define*

$$\Phi^u(A) \stackrel{\text{def}}{=} \text{Tr}(uI - A)^{-1} = \sum_i \frac{1}{u - \lambda_i} \quad \text{(upper potential)},$$

$$\Phi_l(A) \stackrel{\text{def}}{=} \text{Tr}(A - lI)^{-1} = \sum_i \frac{1}{\lambda_i - l} \quad \text{(lower potential)}.$$

As long as $A \prec uI$ and $A \succ lI$ (i.e., $\lambda_{\max}(A) < u$ and $\lambda_{\min}(A) > l$), these potential functions measure how far the eigenvalues of A are from the barriers u and l. In particular, they blow up as any eigenvalue approaches a barrier, since then $uI - A$ (or $A - lI$) approaches a singular matrix. Their strength lies in the fact that they reflect the locations of all the eigenvalues simultaneously: for instance, $\Phi^u(A) \le 1$ implies that no λ_i is within distance one of u, no two λ_i's are at distance 2, no k are at distance k, and so on. In terms of the physical model, the upper potential $\Phi^u(A)$ is equal to the total repulsion of the eigenvalues of A from the upper barrier u, while $\Phi_l(A)$ is the analogous quantity for the lower barrier.

To prove the theorem, we will build the sum $\sum_i s_i \mathbf{v}_i \mathbf{v}_i^T$ iteratively, adding one vector at a time. Specifically, we will construct a sequence of matrices

$$0 = A^{(0)}, A^{(1)}, \ldots, A^{(Q)}$$

along with positive constants[4] u_0, l_0, δ_U, δ_L, ϵ_U, and ϵ_L, which satisfy the following conditions:

(a) Initially, the barriers are at $u = u_0$ and $l = l_0$ and the potentials are

$$\Phi^{u_0}(A^{(0)}) = \epsilon_U \quad \text{and} \quad \Phi_{l_0}(A^{(0)}) = \epsilon_L.$$

(b) Each matrix is obtained by a rank-one update of the previous one—specifically by adding a positive multiple of an outer product of some $\mathbf{v}_i$:

$$A^{(q+1)} = A^{(q)} + t\mathbf{v}\mathbf{v}^T \quad \text{for some } \mathbf{v} \in \{\mathbf{v}_i\} \text{ and } t \ge 0.$$

(c) If we increment the barriers u and l by δ_U and δ_L, respectively, at each step, then the upper and lower potentials do not increase. For every $q = 0, 1, \ldots, Q$,

$$\Phi^{u+\delta_U}(A^{(q+1)}) \le \Phi^u(A^{(q)}) \le \epsilon_U \quad \text{for } u = u_0 + q\delta_U,$$

$$\Phi_{l+\delta_L}(A^{(q+1)}) \le \Phi_l(A^{(q)}) \le \epsilon_L \quad \text{for } l = l_0 + q\delta_L.$$

(d) No eigenvalue ever jumps across a barrier. For every $q = 0, 1, \ldots, Q$,

$$\lambda_{\max}(A^{(q)}) < u_0 + q\delta_U \quad \text{and} \quad \lambda_{\min}(A^{(q)}) > l_0 + q\delta_L.$$

---

[4]On first reading the paper, we suggest the reader follow the proof with the assignment $\epsilon_U = \epsilon_L = 1$, $u_0 = n$, $l_0 = -n$, $\delta_U = 2$, $\delta_L = 1/3$. This will provide the bound $(6d + 1)/(d - 1)$ and eliminates the need to use Claim 3.6.

To complete the proof we will choose u_0, l_0, δ_U, δ_L, ϵ_U, and ϵ_L so that after $Q = dn$ steps, the condition number of $A^{(Q)}$ is bounded by

$$\frac{\lambda_{\max}(A^{(Q)})}{\lambda_{\min}(A^{(Q)})} \le \frac{u_0 + dn\delta_U}{l_0 + dn\delta_L} = \frac{d + 1 + 2\sqrt{d}}{d + 1 - 2\sqrt{d}}.$$

By construction, $A^{(Q)}$ is a weighted sum of at most dn of the vectors, as desired.

The main technical challenge is to show that conditions (b) and (c) can be satisfied simultaneously—i.e., that there is always a choice of $\mathbf{v}\mathbf{v}^T$ to add to the current matrix which allows us to shift *both* barriers up by a constant without increasing either potential. We achieve this in the following three lemmas.

The first lemma concerns shifting the upper barrier. If we shift u forward to $u + \delta_U$ without changing the matrix A, then the upper potential $\Phi^u(A)$ decreases since the eigenvalues λ_i do not move and u moves away from them. This gives us room to add some multiple of a vector $t\mathbf{v}\mathbf{v}^T$, which will move the λ_i toward l and increase the potential, counteracting the initial decrease due to shifting. The following lemma quantifies exactly how much of a given $\mathbf{v}\mathbf{v}^T$ we can add without increasing the potential beyond its original value before shifting.

LEMMA 3.3 (upper barrier shift). *Suppose that $\lambda_{\max}(A) < u$, and $\mathbf{v}$ is any vector. If*

$$\frac{1}{t} \ge \frac{\mathbf{v}^T((u + \delta_U)I - A)^{-2}\mathbf{v}}{\Phi^u(A) - \Phi^{u+\delta_U}(A)} + \mathbf{v}^T((u + \delta_U)I - A)^{-1}\mathbf{v} \stackrel{\text{def}}{=} U_A(\mathbf{v}),$$

then

$$\Phi^{u+\delta_U}(A + t\mathbf{v}\mathbf{v}^T) \le \Phi^u(A) \quad and \quad \lambda_{\max}(A + t\mathbf{v}\mathbf{v}^T) < u + \delta_U.$$

That is, if we add t times $\mathbf{v}\mathbf{v}^T$ to A and shift the upper barrier by δ_U, then we do not increase the upper potential.

We remark that $U_A(\mathbf{v})$ is linear in the outer product $\mathbf{v}\mathbf{v}^T$.

Proof. Let $u' = u + \delta_U$. By the Sherman–Morrison formula, we can write the updated potential as

$$\Phi^{u+\delta_U}(A + t\mathbf{v}\mathbf{v}^T) = \text{Tr}(u'I - A - t\mathbf{v}\mathbf{v}^T)^{-1}$$

$$= \text{Tr}\left((u'I - A)^{-1} + \frac{t(u'I - A)^{-1}\mathbf{v}\mathbf{v}^T(u'I - A)^{-1}}{1 - t\mathbf{v}^T(u'I - A)^{-1}\mathbf{v}}\right)$$

$$= \text{Tr}(u'I - A)^{-1} + \frac{t\text{Tr}(\mathbf{v}^T(u'I - A)^{-1}(u'I - A)^{-1}\mathbf{v})}{1 - t\mathbf{v}^T(u'I - A)^{-1}\mathbf{v}}$$

$$\text{since Tr is linear and } \text{Tr}(XY) = \text{Tr}(YX)$$

$$= \Phi^{u+\delta_U}(A) + \frac{t\mathbf{v}^T(u'I - A)^{-2}\mathbf{v}}{1 - t\mathbf{v}^T(u'I - A)^{-1}\mathbf{v}}$$

$$= \Phi^u(A) - (\Phi^u(A) - \Phi^{u+\delta_U}(A)) + \frac{\mathbf{v}^T(u'I - A)^{-2}\mathbf{v}}{1/t - \mathbf{v}^T(u'I - A)^{-1}\mathbf{v}}.$$

As $U_A(\mathbf{v}) > \mathbf{v}^T(u'I - A)^{-1}\mathbf{v}$, the last term is finite for $1/t \ge U_A(\mathbf{v})$. By now, substituting any $1/t \ge U_A(\mathbf{v})$, we find that $\Phi^{u+\delta_U}(A + t\mathbf{v}\mathbf{v}^T) \le \Phi^u(A)$. This also tells us that $\lambda_{\max}(A + t\mathbf{v}\mathbf{v}^T) < u + \delta_U$, because if this were not the case, then there would be some positive $t' \le t$ for which $\lambda_{\max}(A + t'\mathbf{v}\mathbf{v}^T) = u + \delta_U$. But, at such a t', $\Phi^{u+\delta_U}(A + t'\mathbf{v}\mathbf{v}^T)$ would blow up, and we have thus established that it is finite. $\quad\square$

The second lemma is about shifting the lower barrier. Here, shifting l forward to $l + \delta_L$ while keeping A fixed has the opposite effect—it increases the lower potential $\Phi_l(A)$ since the barrier l moves toward the eigenvalues λ_i. Adding a multiple of a vector $t\mathbf{v}\mathbf{v}^T$ will move the λ_i forward and away from the barrier, decreasing the potential. Here, we quantify exactly how much of a given $\mathbf{v}\mathbf{v}^T$ we need to add to compensate for the initial increase from shifting l, and return the potential to its original value before the shift.

LEMMA 3.4 (lower barrier shift). *Suppose that $\lambda_{\min}(A) > l$, $\Phi_l(A) \leq 1/\delta_L$, and* $\mathbf{v}$ *is any vector. If*

$$0 < \frac{1}{t} \leq \frac{\mathbf{v}^T(A - (l + \delta_L)I)^{-2}\mathbf{v}}{\Phi_{l+\delta_L}(A) - \Phi_l(A)} - \mathbf{v}^T(A - (l + \delta_L)I)^{-1}\mathbf{v} \stackrel{\text{def}}{=} L_A(\mathbf{v}),$$

then

$$\Phi_{l+\delta_L}(A + t\mathbf{v}\mathbf{v}^T) \leq \Phi_l(A) \quad and \quad \lambda_{\min}(A + t\mathbf{v}\mathbf{v}^T) > l + \delta_L.$$

That is, if we add t times $\mathbf{v}\mathbf{v}^T$ to A and shift the lower barrier by δ_L, then we do not increase the lower potential.

Proof. First, observe that $\lambda_{\min}(A) > l$ and $\Phi_l(A) \leq 1/\delta_L$ imply that $\lambda_{\min}(A) > l + \delta_L$. So, for every $t > 0$, $\lambda_{\min}(A + t\mathbf{v}\mathbf{v}^T) > l + \delta_L$.

Now proceed as in the proof for the upper potential. Let $l' = l + \delta_L$. By Sherman–Morrison, we have

$$\Phi_{l+\delta_L}(A + t\mathbf{v}\mathbf{v}^T) = \text{Tr}(A + t\mathbf{v}\mathbf{v}^T - l'I)^{-1}$$
$$= \text{Tr}\left((A - l'I)^{-1} - \frac{t(A - l'I)^{-1}\mathbf{v}\mathbf{v}^T(A - l'I)^{-1}}{1 + t\mathbf{v}^T(A - l')^{-1}\mathbf{v}}\right)$$
$$= \text{Tr}(A - l'I)^{-1} - \frac{t\text{Tr}(\mathbf{v}^T(A - l'I)^{-1}(A - l'I)^{-1}\mathbf{v})}{1 + t\mathbf{v}^T(A - l'I)^{-1}\mathbf{v}}$$
$$= \Phi_{l+\delta_L}(A) - \frac{t\mathbf{v}^T(A - l'I)^{-2}\mathbf{v}}{1 + t\mathbf{v}^T(A - l'I)^{-1}\mathbf{v}}$$
$$= \Phi_l(A) + (\Phi_{l+\delta_L}(A) - \Phi_l(A)) - \frac{\mathbf{v}^T(A - l'I)^{-2}\mathbf{v}}{1/t + \mathbf{v}^T(A - l'I)^{-1}\mathbf{v}}.$$

Rearranging shows that $\Phi_{l+\delta_L}(A + t\mathbf{v}\mathbf{v}^T) \leq \Phi_l(A)$ when $1/t \leq L_A(\mathbf{v})$. $\quad\square$

The third lemma identifies the conditions under which we can find a single $t\mathbf{v}\mathbf{v}^T$ which allows us to maintain both potentials while shifting barriers and thereby continue the process. The proof that such a vector exists is by an averaging argument, so this can be seen as the step in which we relate the behavior of actual vectors to the behavior of the expected vector $\mathbf{v}_{\text{avg}}$. Notice that the use of variable weights t, from which the eventual s_i arise, is crucial to this part of the proof.

LEMMA 3.5 (both barriers). *If $\lambda_{\max}(A) < u$, $\lambda_{\min}(A) > l$, $\Phi^u(A) \leq \epsilon_U$, $\Phi_l(A) \leq \epsilon_L$, and ϵ_U, ϵ_L, δ_U, and δ_L satisfy*

$$(3) \qquad\qquad 0 \leq \frac{1}{\delta_U} + \epsilon_U \leq \frac{1}{\delta_L} - \epsilon_L,$$

then there exist an i and positive t for which

$$L_A(\mathbf{v}_i) \geq 1/t \geq U_A(\mathbf{v}_i), \quad \lambda_{\max}(A + t\mathbf{v}_i\mathbf{v}_i^T) < u + \delta_U,$$

$$and \quad \lambda_{\min}(A + t\mathbf{v}_i\mathbf{v}_i^T) > l + \delta_L.$$

Proof. We will show that

$$\sum_i L_A(\mathbf{v}_i) \geq \sum_i U_A(\mathbf{v}_i),$$

from which the claim will follow by Lemmas 3.3 and 3.4. We begin by bounding

$$\sum_i U_A(\mathbf{v}_i) = \frac{\sum_i \mathbf{v}_i^T((u + \delta_U)I - A)^{-2}\mathbf{v}_i}{\Phi^u(A) - \Phi^{u+\delta_U}(A)} + \sum_i \mathbf{v}_i^T((u + \delta_U)I - A)^{-1}\mathbf{v}_i$$

$$= \frac{((u + \delta_U)I - A)^{-2} \bullet (\sum_i \mathbf{v}_i\mathbf{v}_i^T)}{\Phi^u(A) - \Phi^{u+\delta_U}(A)} + ((u + \delta_U)I - A)^{-1} \bullet \left(\sum_i \mathbf{v}_i\mathbf{v}_i^T\right)$$

$$= \frac{\mathrm{Tr}((u + \delta_U)I - A)^{-2}}{\Phi^u(A) - \Phi^{u+\delta_U}(A)} + \mathrm{Tr}((u + \delta_U)I - A)^{-1}$$

$$\text{since} \sum_i \mathbf{v}_i\mathbf{v}_i^T = I \text{ and } X \bullet I = \mathrm{Tr}(X)$$

$$= \frac{\sum_i(u + \delta_U - \lambda_i)^{-2}}{\sum_i(u - \lambda_i)^{-1} - \sum_i(u + \delta_U - \lambda_i)^{-1}} + \Phi^{u+\delta_U}(A)$$

$$= \frac{\sum_i(u + \delta_U - \lambda_i)^{-2}}{\delta_U \sum_i(u - \lambda_i)^{-1}(u + \delta_U - \lambda_i)^{-1}} + \Phi^{u+\delta_U}(A)$$

$$\leq 1/\delta_U + \Phi^{u+\delta_U}(A)$$

$$\text{as} \sum_i(u - \lambda_i)^{-1}(u + \delta_U - \lambda_i)^{-1} \geq \sum_i(u + \delta_U - \lambda_i)^{-2}$$

$$\leq 1/\delta_U + \Phi^u(A) \leq 1/\delta_U + \epsilon_U.$$

On the other hand, we have

$$\sum_i L_A(\mathbf{v}_i) = \frac{\sum_i \mathbf{v}_i^T(A - (l + \delta_L))^{-2}\mathbf{v}_i}{\Phi_{l+\delta_L}(A) - \Phi_l(A)} - \sum_i \mathbf{v}_i^T(A - (l + \delta_L)I)^{-1}\mathbf{v}_i$$

$$= \frac{(A - (l + \delta_L)I)^{-2} \bullet (\sum_i \mathbf{v}_i\mathbf{v}_i^T)}{\Phi_{l+\delta_L}(A) - \Phi_l(A)} - (A - (l + \delta_L)I)^{-1} \bullet \left(\sum_i \mathbf{v}_i\mathbf{v}_i^T\right)$$

$$= \frac{\mathrm{Tr}(A - (l + \delta_L)I)^{-2}}{\Phi_{l+\delta_L}(A) - \Phi_l(A)} - \mathrm{Tr}(A - (l + \delta_L)I)^{-1}$$

$$\text{since} \sum_i \mathbf{v}_i\mathbf{v}_i^T = I \text{ and } X \bullet I = \mathrm{Tr}(X)$$

$$= \frac{\sum_i(\lambda_i - l - \delta_L)^{-2}}{\sum_i(\lambda_i - l - \delta_L)^{-1} - \sum_i(\lambda_i - l)^{-1}} - \sum_i(\lambda_i - l - \delta_L)^{-1}$$

$$\geq 1/\delta_L - \sum_i(\lambda_i - l)^{-1} = 1/\delta_L - \epsilon_L$$

by Claim 3.6.

Putting these together, we find that

$$\sum_i U_A(\mathbf{v}_i) \leq \frac{1}{\delta_U} + \epsilon_U \leq \frac{1}{\delta_L} - \epsilon_L \leq \sum_i L_A(\mathbf{v}_i),$$

as desired. $\square$

CLAIM 3.6. *If $\lambda_i > l$ for all i, $0 \leq \sum_i (\lambda_i - l)^{-1} \leq \epsilon_L$, and $1/\delta_L - \epsilon_L \geq 0$, then*

$$\frac{\sum_i (\lambda_i - l - \delta_L)^{-2}}{\sum_i (\lambda_i - l - \delta_L)^{-1} - \sum_i (\lambda_i - l)^{-1}} - \sum_i \frac{1}{\lambda_i - l - \delta_L}$$

(4)
$$\geq \frac{1}{\delta_L} - \sum_i \frac{1}{\lambda_i - l}.$$

Proof. We have

$$\delta_L \leq 1/\epsilon_L \leq \lambda_i - l$$

for every i. So, the denominator of the left-most term on the left-hand side is positive, and the claimed inequality is equivalent to

$$\sum_i (\lambda_i - l - \delta_L)^{-2}$$

$$\geq \left(\sum_i \frac{1}{\lambda_i - l - \delta_L} - \sum_i \frac{1}{\lambda_i - l} \right) \left(\frac{1}{\delta_L} + \sum_i \frac{1}{\lambda_i - l - \delta_L} - \sum_i \frac{1}{\lambda_i - l} \right)$$

$$= \left(\delta_L \sum_i \frac{1}{(\lambda_i - l - \delta_L)(\lambda_i - l)} \right) \left(\frac{1}{\delta_L} + \delta_L \sum_i \frac{1}{(\lambda_i - l - \delta_L)(\lambda_i - l)} \right)$$

$$= \sum_i \frac{1}{(\lambda_i - l - \delta_L)(\lambda_i - l)} + \left(\delta_L \sum_i \frac{1}{(\lambda_i - l - \delta_L)(\lambda_i - l)} \right)^2,$$

which, by moving the first term on the right-hand side to the left-hand side, is just

$$\delta_L \sum_i \frac{1}{(\lambda_i - l - \delta_L)^2 (\lambda_i - l)} \geq \left(\delta_L \sum_i \frac{1}{(\lambda_i - l - \delta_L)(\lambda_i - l)} \right)^2.$$

By Cauchy–Schwarz,

$$\left(\delta_L \sum_i \frac{1}{(\lambda_i - l - \delta_L)(\lambda_i - l)} \right)^2 \leq \left(\delta_L \sum_i \frac{1}{\lambda_i - l} \right) \left(\delta_L \sum_i \frac{1}{(\lambda_i - l - \delta_L)^2 (\lambda_i - l)} \right)$$

$$\leq (\delta_L \epsilon_L) \left(\delta_L \sum_i \frac{1}{(\lambda_i - l - \delta_L)^2 (\lambda_i - l)} \right)$$

$$\text{since } \sum (\lambda_i - l)^{-1} \leq \epsilon_L$$

$$\leq 1 \left(\delta_L \sum_i \frac{1}{(\lambda_i - l - \delta_L)^2 (\lambda_i - l)} \right)$$

$$\text{since } \frac{1}{\delta_L} - \epsilon_L \geq 0,$$

and so (4) is established. $\square$

Proof of Theorem 3.1. All we need to do now is set $\epsilon_U, \epsilon_L, \delta_U$, and δ_L in a manner that satisfies Lemma 3.5 and gives a good bound on the condition number. Then, we can take $A^{(0)} = 0$ and construct $A^{(q+1)}$ from $A^{(q)}$ by choosing any vector $\mathbf{v}_i$ with

$$L_{A^{(q)}}(\mathbf{v}_i) \geq U_{A^{(q)}}(\mathbf{v}_i)$$

(such a vector is guaranteed to exist by Lemma 3.5) and setting $A^{(q+1)} = A^{(q)} + t\mathbf{v}_i\mathbf{v}_i^T$ for any $t \geq 0$ satisfying

$$L_{A^{(q)}}(\mathbf{v}_i) \geq \frac{1}{t} \geq U_{A^{(q)}}(\mathbf{v}_i).$$

It is sufficient to take

$$\delta_L = 1, \qquad \epsilon_L = \frac{1}{\sqrt{d}}, \qquad l_0 = -n/\epsilon_L,$$

$$\delta_U = \frac{\sqrt{d}+1}{\sqrt{d}-1}, \qquad \epsilon_U = \frac{\sqrt{d}-1}{d+\sqrt{d}}, \qquad u_0 = n/\epsilon_U.$$

We can check that

$$\frac{1}{\delta_U} + \epsilon_U = \frac{\sqrt{d}-1}{\sqrt{d}+1} + \frac{\sqrt{d}-1}{\sqrt{d}(\sqrt{d}+1)} = 1 - \frac{1}{\sqrt{d}} = \frac{1}{\delta_L} - \epsilon_L$$

so that (3) is satisfied.

The initial potentials are $\Phi^{\frac{n}{\epsilon_U}}(0) = \epsilon_U$ and $\Phi_{\frac{n}{\epsilon_L}}(0) = \epsilon_L$. After dn steps, we have

$$\begin{aligned}
\frac{\lambda_{\max}(A^{(dn)})}{\lambda_{\min}(A^{(dn)})} &\leq \frac{n/\epsilon_U + dn\delta_U}{-n/\epsilon_L + dn\delta_L} \\
&= \frac{\frac{d+\sqrt{d}}{\sqrt{d}-1} + d\frac{\sqrt{d}+1}{\sqrt{d}-1}}{d - \sqrt{d}} \\
&= \frac{d + 2\sqrt{d} + 1}{d - 2\sqrt{d} + 1},
\end{aligned}$$

as desired. □

To turn this proof into an algorithm, one must first compute the vectors $\mathbf{v}_i$, which can be done in time $O(n^2m)$. For each iteration of the algorithm, we must compute $((u + \delta_U)I - A)^{-1}$, $((u + \delta_U)I - A)^{-2}$, and the same matrices for the lower potential function. This computation can be performed in time $O(n^3)$. Finally, we can decide which edge to add in each iteration by computing $U_A(\mathbf{v}_i)$ and $L_A(\mathbf{v}_i)$ for each edge, which can be done in time $O(n^2m)$. As we run for dn iterations, the total time of the algorithm is $O(dn^3m)$.

4. Sparsifiers of the complete graph. Let $G = (V, E)$ be the complete graph on n vertices, and let $H = (V, F, w)$ be a weighted graph of average degree d that $(1 + \epsilon)$-approximates G. As $x^T L_G x = n \|x\|^2$ for every x orthogonal to $\mathbf{1}$, it is immediate that every vertex of H has weighed degree between n and $(1 + \epsilon)n$. Thus, one should think of H as being an expander graph in which each edge weight has been multiplied by n/d.

As H is weighted and can be irregular, it may at first seem strange to view it as an expander. However, it may easily be shown to have the properties that define expanders: it has high edge conductance, random walks mix rapidly on H and converge to an almost uniform distribution, and it satisfies the expander mixing property (see [2] or [10, Lemma 2.5]). High edge conductance and rapid mixing would not be so interesting if the weighted degrees were not nearly uniform—for example, the star graph has both of these properties, but the random walk on the

star graph converges to a very nonuniform distribution, and the star does not satisfy the expander mixing property. For the convenience of the reader, we include below a proof demonstrating that H has the expander mixing property below.

LEMMA 4.1. *Let $L_H = (V, E, w)$ be a graph that $(1 + \epsilon)$-approximates L_G, the complete graph on V. Then, for every pair of disjoint sets S and T,*

$$\left| w(S, T) - \left(1 + \frac{\epsilon}{2}\right) |S| |T| \right| \leq n(\epsilon/2) \sqrt{|S| |T|},$$

where $w(S, T)$ denotes the sum of the weights of edges between S and T.

Proof. We have

$$-\frac{\epsilon}{2} L_G \preceq L_H - \left(1 + \frac{\epsilon}{2}\right) L_G \preceq \frac{\epsilon}{2} L_G,$$

so we can write

$$L_H = \left(1 + \frac{\epsilon}{2}\right) L_G + M,$$

where M is a matrix of norm at most $(\epsilon/2) \|L_G\| \leq n\epsilon/2$. Let x be the characteristic vector of S, and let y be the characteristic vector of T. We have

$$-w(S, T) = x^T L_H y.$$

As G is the complete graph and S and T are disjoint, we also know

$$x^T L_G y = -|S| |T|.$$

Thus,

$$x^T L_H y = \left(1 + \frac{\epsilon}{2}\right) x^T L_G y + x^T M y$$

$$= -\left(1 + \frac{\epsilon}{2}\right) |S| |T| + x^T M y.$$

The lemma now follows by observing that

$$x^T M y \leq \|M\| \|x\| \|y\| \leq n(\epsilon/2) \sqrt{|S| |T|}. \qquad \square$$

Using the proof of the lower bound on the spectral gap of Alon and Boppana (see [14]) one can show that a d-regular unweighted graph cannot κ-approximate a complete graph for κ asymptotically better than (2). We conjecture that this bound also holds for weighted graphs of average degree d. Presently, we prove the following weaker result for such graphs.

PROPOSITION 4.2. *Let G be the complete graph on vertex set V, and let $H = (V, E, w)$ be a weighted graph with n vertices and a vertex of degree d. If H κ-approximates G, then*

$$\kappa \geq 1 + \frac{2}{\sqrt{d}} - O\left(\frac{\sqrt{d}}{n}\right).$$

Proof. We use a standard approach. Suppose that H is a κ-approximation of the complete graph. We will construct vectors x^* and y^* orthogonal to the $\mathbf{1}$ vector so that

$$\frac{y^{*T} L_H y^*}{x^{*T} L_H x^*} \frac{\|x^*\|^2}{\|y^*\|^2}$$

is large, and this will give us a lower bound on κ.

Let v_0 be the vertex of degree d, and let its neighbors be $v_1, \ldots, v_d$. Suppose that v_i is connected to v_0 by an edge of weight w_i, and the total weight of the edges between v_i and vertices other than $v_0, v_1, \ldots, v_d$ is δ_i. We begin by considering vectors x and y with

$$
x(u) = \begin{cases} 1 & \text{for } u = v_0, \\ 1/\sqrt{d} & \text{for } u = v_i, \ i \geq 1, \\ 0 & \text{for } u \notin \{v_0, \ldots, v_d\}, \end{cases}
$$

$$
y(u) = \begin{cases} 1 & \text{for } u = v_0, \\ -1/\sqrt{d} & \text{for } u = v_i, \ i \geq 1, \\ 0 & \text{for } u \notin \{v_0, \ldots, v_d\}. \end{cases}
$$

These vectors are not orthogonal to $\mathbf{1}$, but we will take care of that later. It is easy to compute the values taken by the quadratic form at x and y:

$$
x^T L_H x = \sum_{i=1}^{d} w_i (1 - 1/\sqrt{d})^2 + \sum_{i=1}^{d} \delta_i (1/\sqrt{d} - 0)^2
$$

$$
= \sum_{i=1}^{d} w_i + \sum_{i=1}^{d} (\delta_i + w_i)/d - 2 \sum_{i=1}^{d} w_i/\sqrt{d}
$$

and

$$
y^T L_H y = \sum_{i=1}^{d} w_i (1 + 1/\sqrt{d})^2 + \sum_{i=1}^{d} \delta_i (-1/\sqrt{d} - 0)^2
$$

$$
= \sum_{i=1}^{d} w_i + \sum_{i=1}^{d} (\delta_i + w_i)/d + 2 \sum_{i=1}^{d} w_i/\sqrt{d}.
$$

The ratio in question is thus

$$
\frac{y^T L_H y}{x^T L_H x} = \frac{\sum_i w_i + \sum_i (\delta_i + w_i)/d + 2 \sum_i w_i/\sqrt{d}}{\sum_i w_i + \sum_i (\delta_i + w_i)/d - 2 \sum_i w_i/\sqrt{d}}
$$

$$
= \frac{1 + \frac{1}{\sqrt{d}} \frac{2 \sum_i w_i}{\sum_i w_i + \sum_i (\delta_i + w_i)/d}}{1 - \frac{1}{\sqrt{d}} \frac{2 \sum_i w_i}{\sum_i w_i + \sum_i (\delta_i + w_i)/d}}.
$$

Since H is a κ-approximation, all weighted degrees must lie between n and $n\kappa$, which gives

$$
\frac{2 \sum_i w_i}{\sum_i w_i + \sum_i (\delta_i + w_i)/d} = \frac{2}{1 + \frac{\sum_i (\delta_i + w_i)/d}{\sum_i w_i}} \geq \frac{2}{1 + \kappa}.
$$

Therefore,

$$
(5) \qquad \frac{y^T L_H y}{x^T L_H x} \geq \frac{1 + \frac{1}{\sqrt{d}} \frac{2}{1+\kappa}}{1 - \frac{1}{\sqrt{d}} \frac{2}{1+\kappa}}.
$$

Let x^* and y^* be the projections of x and y, respectively, orthogonal to the $\mathbf{1}$ vector. Then

$$\|x^*\|^2 = \|x\|^2 - \langle x, \mathbf{1}/\sqrt{n}\rangle^2 = 2 - \frac{(1+\sqrt{d})^2}{n}$$

and

$$\|y^*\|^2 = \|y\|^2 - \langle y, \mathbf{1}/\sqrt{n}\rangle^2 = 2 - \frac{(1-\sqrt{d})^2}{n}$$

so that as $n \to \infty$,

$$(6) \qquad \frac{\|x^*\|^2}{\|y^*\|^2} = 1 - O\left(\frac{\sqrt{d}}{n}\right).$$

Combining (5) and (6), we conclude that, asymptotically,

$$\frac{y^{*T} L_H y^*}{x^{*T} L_H x^*} \frac{\|x^*\|^2}{\|y^*\|^2} \geq \frac{1 + \frac{1}{\sqrt{d}}\frac{2}{1+\kappa}}{1 - \frac{1}{\sqrt{d}}\frac{2}{1+\kappa}} \left(1 - O\left(\frac{\sqrt{d}}{n}\right)\right).$$

But by our assumption the left-hand side is at most κ, so we have

$$\kappa \geq \frac{1 + \frac{1}{\sqrt{d}}\frac{2}{1+\kappa}}{1 - \frac{1}{\sqrt{d}}\frac{2}{1+\kappa}} \left(1 - O\left(\frac{\sqrt{d}}{n}\right)\right),$$

which on rearranging gives

$$\kappa \geq 1 + \frac{2}{\sqrt{d}} - O\left(\frac{\sqrt{d}}{n}\right)$$

as desired.   □

5. Conclusion. We conclude by drawing a connection between Theorem 3.1 and an outstanding open problem in mathematics, the Kadison–Singer conjecture. This conjecture, which dates back to 1959, is equivalent to the well-known paving conjecture [1, 5] as well as to a stronger form of the restricted invertibility theorem of Bourgain and Tzafriri [4, 5]. The following formulation is due to Weaver [20].

CONJECTURE 5.1.   *There are universal constants $\epsilon > 0, \delta > 0$, and $r \in \mathbb{N}$ for which the following statement holds. If $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{R}^n$ satisfy $\|\mathbf{v}_i\| \leq \delta$ for all i and*

$$\sum_{i \leq m} \mathbf{v}_i \mathbf{v}_i^T = I,$$

then there is a partition $X_1, \ldots X_r$ of $\{1, \ldots, m\}$ for which

$$\left\| \sum_{i \in X_j} \mathbf{v}_i \mathbf{v}_i^T \right\| \leq 1 - \epsilon$$

for every $j = 1, \ldots, r$.

Suppose that we had a version of Theorem 3.1 which, assuming $\|\mathbf{v}_i\| \leq \delta$, guaranteed that the scalars s_i were all either 0 or some constant $\beta > 0$ and gave a constant approximation factor $\kappa < \beta$. Then we would have

$$I \preceq \beta \sum_{i \in S} \mathbf{v}_i \mathbf{v}_i^T \preceq \kappa \cdot I$$

for $S = \{i : s_i \neq 0\}$, yielding a proof of Conjecture 5.1 with $r = 2$ and $\epsilon = \min\{1 - \frac{\kappa}{\beta}, \frac{1}{\beta}\}$ since

$$\left\| \sum_{i \in S} \mathbf{v}_i \mathbf{v}_i^T \right\| \leq \frac{\kappa}{\beta} \leq 1 - \epsilon \quad \text{and}$$

$$\left\| \sum_{i \in \overline{S}} \mathbf{v}_i \mathbf{v}_i^T \right\| = 1 - \lambda_{\min} \left(\sum_{i \in S} \mathbf{v}_i \mathbf{v}_i^T \right) \leq 1 - \frac{1}{\beta} \leq 1 - \epsilon.$$

As a special case, such a theorem would also imply the existence of *unweighted* sparsifiers for the complete graph and other (sufficiently dense) edge-transitive graphs. It is also worth noting that the $\|\mathbf{v}_i\| \leq \delta$ condition when applied to vectors $\{\Pi_e\}_{e \in E}$ arising from a graph simply means that the effective resistances of all edges are bounded; thus, we would be able to conclude that any graph with sufficiently small resistances can be split into two graphs that approximate it spectrally.

REFERENCES

[1] C. A. AKEMANN AND J. ANDERSON, *Lyapunov theorems for operator algebras*, Mem. Amer. Math. Soc., 94 (1991), 458.

[2] N. ALON AND F. CHUNG, *Explicit construction of linear sized tolerant networks*, Discrete Math., 72 (1988), pp. 15–19.

[3] A. A. BENCZÚR AND D. R. KARGER, *Approximating s-t minimum cuts in $O(n^2)$ time*, in Proceedings of the 28th Annual Symposium on the Theory of Computing (STOC '96), 1996, pp. 47–55.

[4] J. BOURGAIN AND L. TZAFRIRI, *On a problem of Kadison and Singer*, J. Reine Angew. Math., 420 (1991), pp. 1–43.

[5] P. G. CASAZZA AND J. C. TREMAIN, *The Kadison-Singer problem in mathematics and engineering*, Proc. Natl. Acad. Sci. USA, 103 (2006), pp. 2032–2039.

[6] P. CHEW, *There is a planar graph almost as good as the complete graph*, in Proceedings of the Second Annual Symposium on Computational Geometry (SoCG '86), 1986, pp. 169–177.

[7] H. DETTE AND W. J. STUDDEN, *Some new asymptotic properties for the zeros of Jacobi, Laguerre, and Hermite polynomials*, Constr. Approx., 11 (1995), pp. 227–238.

[8] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, MD, 1996.

[9] N. GOYAL, L. RADEMACHER, AND S. VEMPALA, *Expanders via random spanning trees*, in Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '09), 2009, pp. 576–585.

[10] S. HOORY, N. LINIAL, AND A. WIGDERSON, *Expander graphs and their applications*, Bull. Amer. Math. Soc., 43 (2006), pp. 439–561.

[11] R. KHANDEKAR, S. RAO, AND U. VAZIRANI, *Graph partitioning using single commodity flows*, in Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC '06), 2006, pp. 385–390.

[12] A. LUBOTZKY, R. PHILLIPS, AND P. SARNAK, *Ramanujan graphs*, Combinatorica, 8 (1988), pp. 261–277.

[13] G. A. MARGULIS, *Explicit group theoretical constructions of combinatorial schemes and their application to the design of expanders and concentrators*, Probl. Inf. Transm., 24 (1988), pp. 39–46.

[14] A. NILLI, *On the second eigenvalue of a graph*, Discrete Math., 91 (1991), pp. 207–210.

[15] M. RUDELSON, *Random vectors in the isotropic position*, J. Funct. Anal., 163 (1999), pp. 60–72.

[16] D. A. SPIELMAN AND N. SRIVASTAVA, *Graph sparsification by effective resistances*, in Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC '08), 2008, pp. 563–568.

[17] D. A. SPIELMAN AND S.-H. TENG, *Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems*, in Proceedings of the 36th Annual ACM Symposium on Theory of Computing (STOC '04), 2004, pp. 81–90.

[18] D. A. SPIELMAN AND S.-H. TENG, *Nearly-Linear Time Algorithms for Preconditioning and Solving Symmetric, Diagonally Dominant Linear Systems*, CoRR, abs/cs/0607105, 2008; available online at http://www.arxiv.org/abs/cs.NA/0607105.

[19] D. A. SPIELMAN AND S.-H. TENG, *Spectral Sparsification of Graphs*, CoRR, abs/0808.4134, 2008; available online at http://arxiv.org/abs/0808.4134.

[20] N. WEAVER, *The Kadison-Singer problem in discrepancy theory*, Discrete Math., 278 (2004), pp. 227–239.